DOCUMENT RESUME

ED 193 281 TM 800 596

AUTHOR	Everett, Bruce L.; McLaughlin, Donald H.
TITLE	An Exploratory Study of Barriers to Usage of Large Scale Data Bases such as Project TALENT.
INSTITUTION	American Institutes for Research in the Behavioral Sciences, Palo Alto, Calif.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
PUB DATE	Nov 77
GRANT	NIE-G-74-0003
NOTE	35p.
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Data Bases; Data Processing; Educational Research; *Information Dissemination; *Information Utilization; *Researchers; Social Sciences; Surveys; *User Satisfaction (Information); *Use Studies
IDENTIFIERS	*Project TALENT

ABSTRACT

To assess the potential user population for Project TALENT data and to identify barriers to usage of the data for secondary analysis, a small telephone survey of leading large-scale survey researchers was undertaken. The survey was designed to elicit suggestions for strategies for reducing barriers to use of the data. Four levels of barrier were considered: unawareness of the existence of the data base, negative attitudes about secondary analysis, specific difficulties or deficiencies of the data base, and cost of data processing. In the case of Project TALENT, most researchers were aware of its existence, but not of the scope of information contained in the data base; half expressed negative feelings about secondary analysis in general. Most respondents felt that useful data were present and expressed interest in future use. For established researchers, costs at the level required for use of Project TALENT data were not perceived as a barrier. Recommendations obtained from the respondents for improving usage matched the steps taken by the TALENT staff, although some areas were identified where additional effort was needed. (Author/GK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED193281

AN EXPLORATORY STUDY OF BARRIERS TO USAGE OF LARGE-SCALE DATA BASES SUCH AS PROJECT TALENT

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Bruce E. Everett

Donald H. McLaughlin

Supported by
The National Institute of Education

TM 800596

American Institutes for Research November, 1977

AN EXPLORATORY STUDY OF BARRIERS TO USAGE OF LARGE-SCALE DATA BASES SUCH AS PROJECT TALENT

Bruce E. Everett

Donald H. McLaughlin

The research reported here was carried out with funds provided through grant #NIE-G-74-0003 from the Basic Skills Group of the National Institute of Education. The content does not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement should be inferred.

3

American Institutes for Research November, 1977

Table of Contents

	Page
Abstract	iii
Introduction	1
Background Information on the Project TALENT Data Bank	2
Previous Research on the Use of Large Scale Data Bases	4
Focus of the Current Report	6
Methods	8
Sample	8
Instruments	8
Procedures	9
Results	
Respondents' Awareness of Project TALENT	10
Barriers to Use of a Data Base Mentioned by Respondents	10
Strategies Suggested by Respondents for Increasing Use of a Data Base	17
Funding of Research Using Large Scale Data Bases	19
Summary and Recommendations	20
Bibliography	27
Appendix A - Data User Interview Guide	29
Appendix B - Funding Source Interview Guide	31

Abstract

In order to assess the potential user population for Project TALENT data and to identify barriers to usage of the data for secondary analysis, a small telephone survey of leading large-scale survey researchers was undertaken. Because of the dearth of relevant findings, the survey was designed to shed light on barriers to secondary use of large data bases in general as well as on barriers to use of TALENT data. The survey was also designed to elicit suggestions for strategies for reducing barriers to use of the data.

Four levels of barrier were considered: unawareness of the existence of the data base, negative attitudes about secondary analysis, specific difficulties or deficiencies of the data base, and cost of data processing. In the case of Project TALENT, most researchers were aware of its existence, but few had a clear idea of the scope of information contained in the data base. Half of the respondents expressed some form of negative feeling about secondary analysis in general, and a folklore myth that the TALENT data are severely biased was also uncovered. When the data base was described to the respondents, most felt that useful data were present and many expressed interest in future use of the data. For established researchers, costs at the level required for use of Project TALENT data were not perceived as a barrier, although they might be for beginning researchers.

Recommendations obtained from the respondents for improving usage generally matched the steps recently taken by the TALENT staff (see Chapter 9 of this report), although some areas were identified where additional effort was needed (e.g., publishing more articles based on Project TALENT). Finally, continuation of activities developed in this survey is desirable both to refine our knowledge about secondary data usage patterns and, as a side effect, to heighten interest in secondary analysis of TALENT data.

AN EXPLORATORY STUDY OF BARRIERS TO USAGE OF LARGE-SCALE DATA BASES SUCH AS PROJECT TALENT

Introduction

Large-scale data bases in the social and behavioral sciences provide tremendous advantages, both in terms of cost and in terms of the opportunity for addressing important research questions that are normally beyond the reach of most investigators. As such, they should be viewed as a data resource--a means of acquiring key information at a fraction of the cost of gathering new data. The importance of secondary analysis as a research tool has been discussed by several authors (Koehler, 1977; Glass, 1976; Hyman, 1972), but the methods of matching researchers to data have not yet been thoroughly explored. The American Institutes for Research (AIR) feels a responsibility for taking positive action to promote secondary analyses of Project TALENT, a major data base that it maintains, and toward this end the TALENT staff has been engaged in a serious effort to make Project TALENT data more accessible to researchers.

This report describes a study of the opinions and attitudes of professional survey researchers concerning the potential for utilizing large-scale data bases such as Project TALENT. The intent of the study was to identify the barriers that influence secondary analyses and make specific recommendations to encourage further use of existing data resources.

If a data resource is to be utilized as a data bank and be readily available to other researchers, then there are clearly responsibilities for its managers. These include the maintenance of a qualified staff of consultants, the dissemination of technical information, the cataloging of uses of the data, the review of user needs and problems, and the periodic upgrading of the contents of the data bank and the services provided (Nasatir, 1973).

What is less obvious is the extent to which the managers of a data base should be expected to go to encourage the use of their data. Merely making the data available may be professionally acceptable, but this does not solve the problem; without extensive publicity and dissemination efforts, very few other researchers are likely to use a data base. Unfortunately, encouraging outside researchers to make use of a large-scale data base is no easy task: as will be discussed in this report, there are definite barriers to the use of large-scale data bases that have to be overcome, both by the prospective users and by the data base managers.

The problem of data dissemination is particularly difficult in the social sciences, when attention to individual respondents' rights to privacy must be given highest priority. Merely deleting a respondent's name from the data record does not preclude his/her identification in terms of a unique combination of experiences (e.g., birth date, college major, occupation, number and age of children, military experience). Responsible dissemination must recognize this problem and deal with it. In the case of Project TALENT, data have been released only to researchers for their own use and only after they have agreed in writing not to use the data in any way that would identify individual respondents.

This report addresses the issue of how best to promote widespread use of large-scale data bases, examining the perceived and actual facilitators and barriers to the use of a particular data base. The report focuses on the Project TALENT data base, a large-scale data resource that is potentially useful to a broad spectrum of researchers.

Background Information on the Project TALENT Data Bank

Project TALENT is an ongoing nationwide study of some 400,000 American men and women who were in high school in 1960. It has been supported through the years first largely by the U.S. Office of Education and, more recently, wholly by the National Institute of Education. The TALENT data base includes measures of the cognitive skills, interests, plans, family backgrounds, and current activities of the original 1960 sample, plus data collected one year after high school on over

190,000 of its participants, data collected five years after high school on over 130,000 of its participants, and data collected eleven years after high school on over 95,000 of its participants. Plans are under way for a seventeen-years-after-high-school follow-up survey that will extend this series of longitudinal studies of the educational, occupational, family and life style history of this representative cross-section of American society.

In order to facilitate the use of TALENT data by the research community, the American Institutes for Research maintains the Project TALENT Data Bank. Data Bank staff are available to consult with researchers to determine their particular needs, to interpret them in terms of possible Project TALENT contributions, and to supply tapes or carry out the necessary computer runs and analyses. Preliminary planning for the use of Project TALENT data has been facilitated by publication of a comprehensive Data Bank Handbook which describes the data base, the procedures used for sampling and data collection, and how to locate and specify variables of interest (Wise et al., 197).*

Between 1960 and 1977, over 200 studies, conducted both by TALENT staff and by outside researchers, have utilized Project TALENT data. The data have been used to support Congressional testimony on vocational education, higher education, guidance and counseling, and fertility questions. Studies have been made of the interactions among race, sex, socioeconomic status, and various ability measures, and of their effects on subsequent educational and career attainment. TALENT data have been used to assess the effects of school characteristics, curriculum and guidance opportunities, career preferences, and marital and family history on postsecondary education and career success. As one researcher said, "the best currently available evidence about high schools' effects on their students is found in survey data collected by Project TALENT" (Jencks et al., 1972, p. 89).

---

*The Handbook may be obtained for $6.00 from the American Institutes for Research, P.O. Box 1113, Palo Alto, California 94302. Readers wishing further information regarding Project TALENT or the Project TALENT Data Bank are invited to contact either Dr. Lauress Wise, Deputy Director of Project TALENT, or Dr. Donald McLaughlin, Director of the Project TALENT Data Bank, at the American Institutes for Research.

In addition to Jencks' study of the effects of family and schooling on inequality, several other major studies have made use of Project TALENT data. For example, the recent highly publicized study conducted by the Educational Testing Service concerning the decline in test scores (Turnbull et al., 1977) relied heavily on Project TALENT data. The list of publications by John Flanagan and others on the Project TALENT staff includes dozens of separate references.

Previous Research on the Use of Large-Scale Data Bases

Before presenting our own findings concerning potential barriers and facilitators to the use of large-scale data bases, the results of other similar studies should be mentioned. Ennis (1964) found that only 21% of the data sources used in articles published in 15 major American sociological journals in 1962 had led to other publications. Bell (1970) pointed out that not only do the majority of social scientists avoid secondary analysis of data, but those who do use secondary analysis tend to be situated at the larger colleges and universities. Only one of Bell's 15 respondents stated that he had first learned of a data source through a publication; eight discovered the existence of their data through lists published by data archives; and six found out about the data through casual conversations or correspondence with colleagues. Ten of these 15 said that they were disappointed with the utility of the data once they received it. Bell concluded that prospective users tend to be uninformed about the contents of a secondary data source, they often overestimate the costs of data acquisition, and they may simply not have the skills required for secondary analysis. The barriers to secondary data use that he found were inefficient search techniques in locating data sources, lack of information about the data, and poor quality in data coding.

Babbie (1973) enumerated some of the responsibilities of the original researcher when a data base is released for secondary analysis. First, the original researcher should prepare a methodological report on his study, not only indicating the manner in which the study was conducted but also pointing to special strengths and weaknesses in the data. Second, the original researcher should request copies of reports

prepared from his data in order to review them for inaccuracies and misinterpretations. Finally, the original researcher should challenge any misuses of his data in secondary analyses.

Hyman (1972) provided probably the richest source of information concerning the problems of using an existing data source. He cites lack of training in secondary analysis techniques as a major obstacle, as well as the time and effort involved, lack of awareness of data sources, delays in obtaining data, costs involved, the quality of the data and its documentation, and the absence of key variables. The benefits of secondary analysis that he mentioned are that it economizes on time, money, and personnel, reduces the intrusion into the lives of subjects and respondents, provides a training ground for beginning researchers, allows studies involving changes over time or multinational settings, affords the possibility for multiple replication of survey findings, and compels researchers to think more broadly and abstractly about their work.

Hyman's examples of successful attempts at secondary analysis provide an excellent description of the target population both for this study and for future dissemination efforts concerning Project TALENT data. The characteristics of researchers who are likely to perform secondary analyses, he found, include the possession of varied and broad interests, a tolerance of minor imperfections in a data source, sensitivity to the opportunities that are presented, a wide network of information sources, an understanding of how completely a vein of information can be worked and when to move on to other data, an open mind as to what directions one's inquiries should take, and an ability to see the larger potential of even minor indicators.

Hyman characterized the experienced secondary analyst as a person who is able to cope with the obstacles normally standing in the way of adequate treatment of error, concerned about the ambiguity or invalidity of his indicators as instruments for the measurement of particular variables, and familiar with the often complex designs of survey research. Above all, secondary analysts must avoid being either too rigid

in their data requirements or too unsystematic in their search for new data--"they are purposeful in their search, but still relatively easy to please. Since their pursuit is likely to be rewarded, their affection [for the data] grows. As the relationship with the archive persists, they become more familiar with [its] many charms and the quirks in [its] data, and progressively more skillful in their dealings. Each later encounter is simpler and easier. Thus a fruitful and gratifying relationship [between researcher and data base] develops" (Hyman, 1972, p. 79).

Focus of the Current Report

As part of the Project TALENT activities funded under Grant No. NIE-G-74-0003, the current study was undertaken to determine the factors that inhibit the use of data bases such as Project TALENT. This study adds to the work of Hyman (1972), Nasatir (1973), Babbie (1973) and others who have addressed the problems of access to large-scale data bases and secondary analysis.

The current study focuses specifically on obtaining the informed opinion of experienced survey researchers to address the following questions:

1. How knowledgeable are researchers about Project TALENT and its data base? What are their sources of information?

2. What are the most effective methods for disseminating information about Project TALENT and encouraging the use of its data base?

3. What are the barriers that need to be overcome before researchers will want to use Project TALENT data? Who would be most likely to use the data?

4. What can Project TALENT staff do to increase the use of TALENT data by other researchers?

The remaining sections of this report describe the procedures followed in carrying out this study, the results obtained, and the implications for managers and sponsors of large-scale data bases.

Methods

Sample

Rather than restrict the sample solely to secondary analysts, it was decided that researchers in the social sciences who were experienced in the use of large-scale survey data should be the primary source of respondents. In addition, efforts were made to contact others who were knowledgeable about the use of data banks and the funding of research projects based on existing survey data. Ninety-five potential research respondents were identified from the educational, psychological, sociological, vocational, and demographic literature for the past two years as authors who had recently conducted studies using survey data similar to those available in Project TALENT. Mail and telephone responses were obtained from 50 individual researchers: 18 sociologists, nine survey experts, seven economists, five labor experts, five involved in population studies, three policy researchers, two educators, and one psychologist. Of these, four had previously used TALENT data. In addition, sixteen representatives of Federal funding sources were considered as possible survey respondents, and five of these were contacted by telephone. Thus a total of 55 interviews were conducted.

Instruments

An interview protocol was devised to elicit comments from the respondents regarding the use of Project TALENT and other similar data bases. The open-ended questions were intended merely to guide the conversation in order to capitalize on the personal opinions and suggestions of the respondents. Although the emphasis was on respondents' reactions to Project TALENT, they were also given the option of talking about large-scale data bases in general. Respondents were asked about their previous sources of information about Project TALENT, the nature of their current research, the problems they had encountered in using large-scale data bases such as Project TALENT, the suggestions that they had for increasing the use of Project TALENT data, and the topical areas that they most wanted to see included in the next TALENT survey (see Appendix A for a copy of the interview guide).

A slightly different interview guide was used to obtain comments from the five Federal officials. These respondents were asked to describe their sources of information about Project TALENT, the kinds of survey research efforts that they sponsor, the barriers that prevented their funding research using large-scale surveys, and what they thought Project TALENT could do to increase its relevance to government research needs (see Appendix B for a copy of the Interview Guide).

Procedures

After the pool of respondents had been decided upon, they were sent an introductory letter containing an explanation of the purpose of the study, a brochure describing Project TALENT, and the set of questions to which they were requested to respond. Of the 111 individuals who had been mailed questions, five respondents mailed in their replies, and the rest were contacted by telephone. Three declined to be interviewed, and 53 could not be reached by telephone within the time limits of the survey. Thus responses were obtained from a total of 55 individuals.

Respondents were first asked if they had received the mailed version of the survey and if they had formed some opinions concerning it. (Virtually all of the respondents had received the mailed version, but most had not taken the time to formulate their responses.) They were then asked each of the questions in turn, with occasional prompts inserted to stimulate critical opinions, e.g., "Do you feel costs to be a barrier?--What about support services?" At the end of each telephone interview, notes on the conversation were recorded for later analysis.

After all of the telephone interviews had been completed and transcribed, the responses to each question were consolidated into major response categories to facilitate interpretation of the findings. Especially constructive suggestions and criticisms were also extracted for further action.

Results

Respondents' Awareness of Project TALENT

Table 1 presents the responses to the first questions of interest, whether the respondents had heard of Project TALENT before and what their sources of information about Project TALENT were. As might be expected of a sample of survey researchers, nearly all of the respondents claimed at least to have heard of Project TALENT before. Thus, if the results of this survey generalize to the population, it appears that lack of awareness of Project TALENT is not preventing researchers from making use of the Data Bank, but that mere awareness in itself is not sufficient to lead to use of Project TALENT.

It was difficult for most respondents to describe accurately what their sources of information on Project TALENT were. They could recall having talked to other researchers about TALENT findings or encountering references to TALENT in the research literature, but few could remember a specific reference. Only those who had a personal interest in the TALENT data and who had obtained copies of Project TALENT reports showed any real familiarity with the study. Because the term "talent" has been applied to many endeavors, including the NSF Talent Search, it is possible that some respondents who claimed awareness were mistaken.

Barriers to Use of a Data Base Mentioned by Respondents

Table 2 presents a summary of the comments of the respondents regarding the barriers that they saw to the use of large-scale survey research data like Project TALENT by outside researchers. Their responses have been grouped into five major areas of concern:

1. unfamiliarity with the data base and its contents,
2. attitudes against secondary analysis,
3. difficulties in carrying out secondary analysis,

Table 1

Respondents' Awareness of Project TALENT*

Amount of Information	Respondents	
	N	% of Total (46)*
Had not been aware of Project TALENT	5	(11%)
Generally aware of the project; heard about it through colleagues	30	(67%)
Had read articles about Project TALENT	22	(49%)
Had read TALENT publications	11	(24%)

*Limited to the 46 respondents who had not used Project TALENT data. The four previous TALENT users indicated that in addition to conversations with colleagues and journal articles, they had read Project TALENT publications such as the Data Bank Handbook in the course of their research. Five federal respondents were excluded.

Table 2

Barriers to Use of a Data Base Mentioned by Respondents

N=50

Type of Barrier	Respondents	
	N	%
Unfamiliarity with the data	50	(100%)
Data bases generally	13	(26%)
TALENT in particular	37	(74%)
Attitudes against secondary analysis	25	(50%)
Suspicion about data validity	13	(26%)
Desire to collect own data	17	(34%)
Difficulties in secondary analysis	48	(96%)
Absence of key variables	33	(66%)
Quality of the data	21	(42%)
Nonaccess to raw data	22	(44%)
Nonrepresentativeness of the sample	14	(28%)
Complexity of the data base	7	(14%)
Costs of secondary analysis using large data bases	31	(62%)
Obtaining funding	10	(20%)
Expenses of data acquisition	23	(46%)
Poor services provided by the data bank staff	31	(62%)
Quality of the services	23	(46%)
Quality of the documentation	13	(26%)

4. costs of acquiring information from a data bank, and

5. poor quality of data bank services.

Each of these areas is discussed below.

Unfamiliarity with the data. It appears that the respondents viewed a lack of knowledge about a data base and its contents as the major barrier to outside research. Seventy-four percent of the respondents said or implied that this was the case with Project TALENT, and 26% made a general reference to unfamiliarity as being a barrier to secondary analysis.

Although information about Project TALENT and its Data Bank activities has been published in many places, the level of detail necessary to prepare a researcher to make use of the Project TALENT data base is contained only in the Data Bank Handbook. Given the relatively low level of distribution of the Handbook, it is not surprising that so many respondents did not consider themselves especially familiar with Project TALENT. Even respondents who had once been knowledgeable about TALENT admitted that they were not familiar with the contents of the more recent follow-up surveys. As one respondent remarked, "The problem with TALENT is that you don't have anything like the Coleman Report to point to." His concern was that without a definitive or landmark publication available, it would be difficult for most researchers to see a link between Project TALENT and their own research interests.

Further support for this finding was evident in the reactions of respondents who showed active interest in using Project TALENT data for their own research. Although most had not yet made a direct effort to obtain TALENT data, their interest in doing so increased as they were made aware of the availability of variables pertinent to their research. This indicates that researchers need to feel knowledgeable about the contents of the Project TALENT data base before they will consider it for their own use.

Attitudes toward secondary analysis. The second set of barriers, those dealing with attitudes toward secondary analysis, also appeared to be a substantial barrier to the outside use of a large-scale data base. Fifty percent of the respondents mentioned a concern about the validity of the data or a desire to work directly with one's own data. Several respondents indicated that because they were accustomed to collecting their own data and carrying out their own analyses, the use of an additional data source did not appeal to them. To quote one respondent, the advantage of creating one's own data base is that "I have the data--not analyses someone has run for me. During the course of any one day I am likely to develop new questions, attack the data from new angles, and generally change in response to the data."

The predisposition of many researchers to avoid using someone else's data conflicts with the obvious fact that it is often far more efficient to utilize an existing large-scale data base than to attempt to create a new small one. Although it depends upon the individual's particular research requirements, e.g., whether he absolutely needs a certain set of variables, many researchers do not like to compromise their research design simply to conform to the nature of existing data. There appears to be a feeling of satisfaction in creating and controlling one's own data sources that secondary analysis cannot match. However, if a data base contains information that meets a researcher's unique needs, this barrier would be considerably reduced.

Difficulties in secondary analysis. Nearly all of the respondents pointed out at least one potential difficulty with the conduct of a secondary analysis, especially when using a large-scale data base. The absence of variables essential to a researcher's specific interests was the most frequently mentioned problem, but concern was also expressed about the quality of the data, the inability to access raw data, possible nonrepresentativeness of the sample, and the complexity of the data base. Many of these same problems can also occur in primary analysis, of course, and the implication is that researchers insist on higher standards of quality when they use someone else's data than they demand of their own data.

Most of the respondents had at one time or another performed at least one secondary analysis using a large-scale data base, such as census data, so they tended to be familiar with the problems of such data bases. They did, however, differ considerably in the degree of intensity with which they regarded these potential difficulties as actual barriers to research. The respondents tended to view the process of getting acquainted with a data source and modifying it to meet their needs as a necessary step in secondary analysis; their main concern was that a data base actually contain key research variables and represent the population of interest.

In talking with respondents who were interested in the possibility of using Project TALENT data regarding actual difficulties in using Project TALENT, there were only a few areas where respondents felt that the measures obtained from TALENT participants were inadequate to arouse their research interest. This was partially due to the fact that the respondents were chosen from among those doing research in areas for which Project TALENT might be relevant. Those respondents interested in demographic changes affecting the TALENT samples (mobility, number of children, marital status, etc.) indicated that the inclusion of such items in the 17-year follow-up survey would increase the attractiveness of the data base for them. Also, several of the economists expressed an interest in obtaining more complete information on income and expenditures of the respondents, an addition that is considered for the 17-year follow-up survey.

Much to our surprise, those respondents who were critical of the quality and representativeness of the Project TALENT data base were apparently not aware of the extensive efforts that have been made to maximize the data quality and to ascertain the precise degree of bias attributable to nonresponse on key variables and to attrition effects in general. There seems to be a folklore about Project TALENT that insists that the sample is biased--a charge that must be addressed more vociferously than in the past. Perhaps, such titles as "When is Bias Better?" should be avoided as they relate to TALENT because they tend to reinforce superficial attitudes about the representativeness of TALENT.

Although several technical descriptions of the proper statistical adjustments to the data exist, including within the TALENT Data Bank Handbook, they do not appear to be common knowledge; instead, what seems to be a more prevalent (yet illogical) attitude is that the size and age of the TALENT data base compounds rather than corrects for such problems with the data.

Costs of secondary analysis using large data bases. A fourth set of barriers to the use of a data base is the costs involved. Depending on the perspective of the researcher, this may or may not be a critical factor; as might be expected, older, more experienced researchers saw little personal difficulty in obtaining the funds necessary to cover data processing and analysis, whereas it was generally agreed that beginning researchers would find all but minor costs a definite barrier.

There were a number of complaints about the current costs of analyses using TALENT data, simply because the respondents had their own, less expensive sources of data processing. University-sponsored computer time and consulting services, not to mention access to graduate student labor, make it preferable for many researchers to obtain copies of the raw data and do their own processing and analysis. Therefore, for these respondents providing worktapes, as Project TALENT does, is a more attractive inducement to use the data. On the other hand, two of the respondents without such resources at their disposal said that if they were to use TALENT data, they would prefer to have TALENT staff do as much of the computational work as possible.

Poor services provided by the Data Bank staff. The fifth set of barriers mentioned was with respect to the quality of the services and documentation provided by those responsible for a data base. Potential problems included mistakes in processing, overexpenditure of funds, time delays, excessive paperwork, incomplete answers to questions, and cryptic or erroneous documentation. These barriers were described by the respondents in terms of what they hoped to avoid in using a particular data base, rather than as specific criticisms.

Strategies Suggested by Respondents for Increasing Use of a Data Base

Table 3 lists the suggestions made by the respondents as to what Project TALENT could do to increase the use of its data. The most frequently mentioned suggestion was that TALENT increase dissemination of descriptions of its data base; however, respondents did not show much agreement on the most effective method. Direct mailing to researchers was mentioned, as were selective mailings to heads of departments, data archivists and survey research experts. Recommendations on the size and contents of the literature to be sent ranged from short, general descriptions to a complete data handbook.

One of the problems of disseminating information about a data base is that relatively few researchers would be likely to read a complete description of the data, yet they would not be personally satisfied by a global description. Those respondents actively interested in using TALENT data made it clear that they wanted a full, variable-by-variable description of the data base. One respondent, a data archivist, pointed out that the level of detail desired by a researcher is closely related to the intensity of his or her interest in that kind of data: a corollary of this is that researchers tend to wait until the onset of their research before making extensive inquiries about a data base.

Respondents appeared to be more impressed by person-to-person, researcher-to-researcher contact regarding a data base than they were by written presentations. It was suggested that Project TALENT follow this procedure wherever feasible, such as in presentations and demonstration booths at conventions and seminars about accessing TALENT data, as well as our current policy of free consultation about prospective research.

Another suggestion was to provide Public Use Samples of the Project TALENT data, such as the one already on file with NIE, for the cost of the computer tape containing them and the necessary documentation. Respondents noted that this would (1) make Project TALENT directly accessible to a large number of users, (2) allow users with limited funds or their own computer resources to use the data inexpensively, and (3) allow researchers to do preliminary studies of the data before requesting analyses of the entire sample.

Table 3

Respondents' Suggestions for Increasing the Use of Project TALENT Data

N=50

Suggestion	Respondents	
	N	%
Provide detailed documentation of data samples	31	(62%)
Increase general dissemination of information about Project TALENT	27	(54%)
Contact potential researchers directly	17	(34%)
Make the raw data available to researchers and data archives	15	(30%)
Defend the representativeness and the value of the TALENT data base	9	(18%)
Be more responsive to the data requirements of potential users, especially in the selection of the questions asked in the surveys	5	(10%)
Publish a bibliography and description of previous studies that used TALENT data	4	(8%)
Improve contacts with funding agencies	4	(8%)
Send dissemination packages to data resource authorities (department heads, data archivists, etc.)	3	(6%)

Some respondents said that they would like to find out what studies had already been done using TALENT data so that they could better relate TALENT to their current research interests. The two reasons given for this were to avoid a repetition of an existing study and to get some indication of the range of studies that are possible with TALENT data.

A notable generalization of the suggested approaches to broadening the data bank usage is that most of the suggestions have already been implemented by TALENT staff, in one way or another. Without some clear idea of the costs and effects of the alternatives suggested, the suggestions do not appear to add substantially to our knowledge for making dissemination decisions.

Funding of Research Using Large-Scale Data Bases

The interviews conducted with representatives of Federal agencies indicated that an effort to make such agencies aware of the value of a data base is as important as the need to contact potential users. Agencies tend to be uninformed about data bases that they do not directly sponsor, and they are not likely to include such data bases in their own funding plans. In addition, the current tendency is to fund the creation of data bases through open-bidding contracts and to fund secondary analyses through grants (Chinitz, 1971). This places a burden on prospective users of a data base in that they have to convince the funding agency of not only the value of their research but also the validity of their data source.

The education of government personnel about the Project TALENT data base definitely needs to be improved: as one respondent said, "I can't recall Project TALENT ever being mentioned in our program planning sessions." Although many agencies are receptive to the notion of TALENT-based research as part of their grant programs, there do not appear to be many agencies other than NIE who are actively planning to sponsor such studies through contract awards. This suggests that a careful review of appropriate grant programs, increased contact and consultation with the government officials responsible for the administration of those programs, and communication to prospective users of likely sources of grant support would be the most effective strategies for increasing the likelihood of funding for TALENT-based research.

Summary and Recommendations

This was a small, exploratory study designed to help the American Institutes for Research and the National Institute of Education decide upon an effective strategy for promoting the more widespread use of the Project TALENT data base. As pointed out in the introduction, the Project TALENT data base, like other very large computerized data bases in the social and behavioral sciences, has the potential for supporting research on so many alternative problems that the choice of problems to pursue is not straightforward. It is most efficiently made at least in part by researchers who conceptually "come to the data base" with a significant problem. Only part of the potential can be realized by a single set of researchers with their specific research agenda. Therefore, AIR has maintained the Project TALENT Data Bank service, through which nearly 150 studies have been performed by researchers outside AIR since 1960; and since 1975 NIE, through its Education and Work Division, has provided a public use sample of Project TALENT data for studies relating education and work.

The purpose of the present study was to find out whether there was potential for much more widespread use of Project TALENT data, and if so, to assess the likely effectiveness of alternative plans for expanding TALENT usage while at the same time maintaining the privacy rights of the 400,000 TALENT participants. Was it possible that the vast majority of potential TALENT users had not heard of TALENT? Were those who were aware of TALENT unaware of the breadth of the data within it? Were there misconceptions about weaknesses that could beset a project like TALENT but have been dealt with by TALENT staff? Were there general inhibitions against secondary analysis? Had the TALENT questionnaires omitted critical data elements? Or were costs of data acquisition a significant barrier? While it was not expected that this study would provide definitive answers to all these questions, it was hoped that interviews with a few dozen leading social researchers in various fields would provide a global, if not exact, perspective from which to proceed further.

There were four main results: (1) most large-scale survey researchers were aware of Project TALENT; (2) however, few other than those who had previously used Project TALENT data had more than a vague idea of what was in the data base. (3) There was substantial interest among researchers in using TALENT data once a description was given, and (4) established researchers did not see the costs of data acquisition from Project TALENT as a significant barrier.

There were numerous secondary results. Some researchers harbored a misconception that Project TALENT staff had not dealt adequately with the attrition problem; many suggestions were received for questions to be included in the next Project TALENT follow-up survey to make the data most useful to different individuals; a few researchers would not seriously consider secondary data analysis for themselves; for some researchers the cost of data acquisition, even a few hundred dollars, presented a problem; and for some researchers, the need to become immersed in data before specifying exact analyses meant that they could not become data bank users without a sample of the data to explore.

In some cases, these results have been anticipated in actions taken by AIR and NIE in the recent years of Project TALENT. For example, the need for small public use samples that can be loaned to potential data bank users for initial exploration is now satisfied with the production of a 4,000-case self-weighted sample of respondents to the Project TALENT 11-year follow-up survey. In other cases, the results suggest actions to be taken (or not to be taken).

A process model such as shown in Figure 1 provides a framework for summarizing and applying the results of this study to planning for increased usage. If the frequency of each step in the chain in Figure 1 is increased, the total number of users will increase, but the most efficiency will be achieved by focusing dissemination efforts on the steps now occurring with lowest frequency. While this model is overly simplistic, it does provide a context for interpreting the percentages presented in Tables 1 and 2.

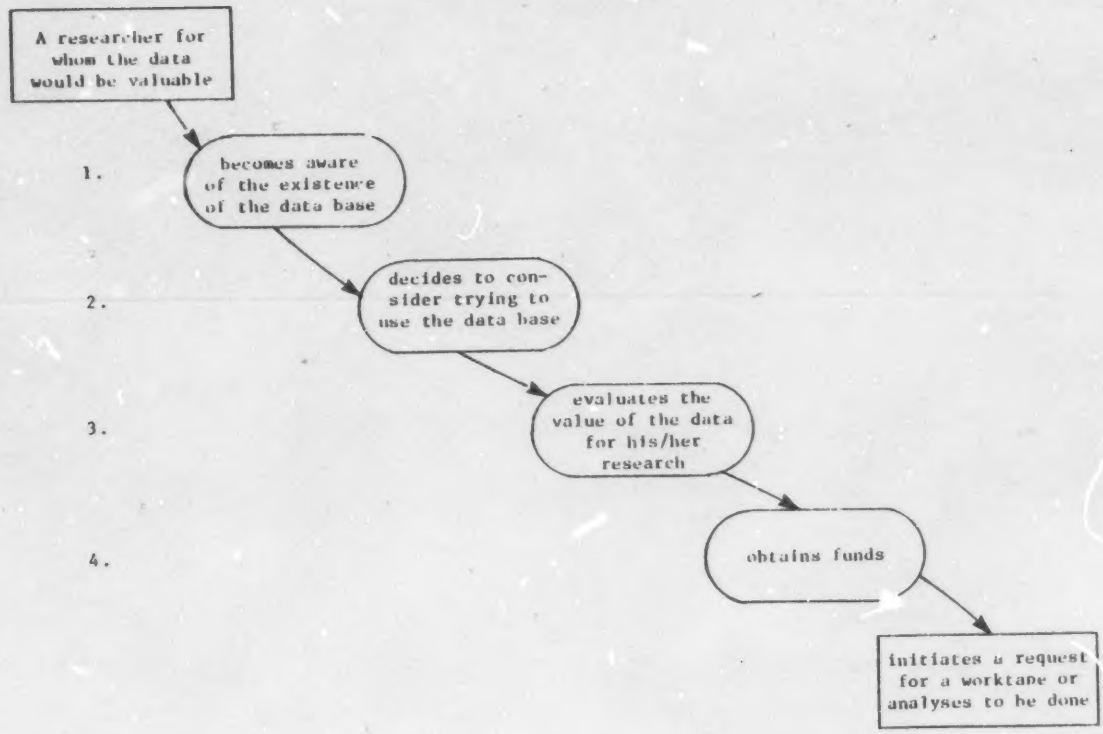
A researcher for whom the data would be valuable
1.
becomes aware of the existence of the data base
2.
decides to consider trying to use the data base
3.
evaluates the value of the data for his/her research
4.
obtains funds
initiates a request for a worktape or analyses to be done

Figure 1. Processes leading to Data Bank usage

In order to decide whether reaching 10% or 30% or 70% of the population is a proper goal for some dissemination activities, we consider a crude example. Using the four-step model in Figure 1, suppose that there were 10,000 researchers for whom the data were potentially valuable. Now, if the conditional probability of each step's occurrence were .10 (e.g., 10% of those who reached the conclusion that the data were valuable for their research were then able to obtain funds), this would lead to a single user ($1 = 10{,}000 \times (.10)^4$). In the past decade, approximately 100 researchers have become users of Project TALENT data. Using the estimate of 10,000 potential users (we need a better estimate in order to refine this model), we arrive at an estimate of about .30 for each of the four steps, on the average. Thus, steps for which we estimate the probability already to be higher than .30 are not as important as barriers as are other steps.

Concerning the first step, in the process model the general awareness of Project TALENT among potential research users was quite high; therefore, unless there are other potential users not tapped by this survey, efforts aimed at merely broadening awareness of Project TALENT need not be expanded. That does not mean they should be discontinued, of course, because general awareness is a function of the appearance of articles and reports and of oral presentations at conventions and would soon fade in the absence of these.

The second step, forming a commitment to consider using TALENT data, is much more problematic. There are three barriers to be overcome in convincing a potential user to look carefully at the data and to apply the data to his/her problem if appropriate. First, there is the hesitancy to use somebody else's data and "merely" perform secondary analyses. Although this survey focused on researchers who had used survey data not unlike the contents of Project TALENT's data base, there were nevertheless quite a few responses that indicated an unwillingness to consider using the data. That attitude is not easily changed; however, until it is, a large reservoir of potential users will remain untapped. One possible tactic to overcome this barrier is to select, say, a half dozen of the most promising areas and devote about two

person-years to production of articles for leading research journals that use TALENT data to address problems of high current relevance. These articles could be written to include persuasive arguments for the importance, feasibility, and efficiency of secondary analyses.

In addition to hesitancy to perform secondary analyses, there is the general problem of inertia. Productive researchers have ongoing research agendas that may not allow time to consider alternatives such as Project TALENT. This is one reason that a large proportion of the recent Project TALENT data bank users have been graduate students. A corollary has been that projects requiring substantial work for a small budget have been the rule, not the exception. The American Institutes for Research and the staff of the Project TALENT Data Bank have donated substantial efforts to helping these graduate student users, but without an independent source of support such efforts cannot be expanded and, in fact, will probably have to be discontinued.

The problem of inertia may be best overcome by creating "small investment" steps which busy researchers will be willing to undertake and which will provide a compelling argument for further interest in data bank usage. One step taken by Project TALENT staff in this area has been production of the report, The American Citizen: 11 Years After High School, in which potential researcher users can quickly find the response distributions to follow-up questions that might form the basis for a study and at the same time assess relations of these responses to variables measured in high school. Selective mailing of this document to researchers previously contacted by telephone and expressing interest in finding out about TALENT research results could prove fruitful.

The third attitudinal problem, in addition to hesitancy to perform secondary analysis and inertia, is mistrust of the validity of the data. For long-term longitudinal analysis the most prevalent grounds for mistrust concern attrition. There appears to be a consensus in some circles that Project TALENT staff have not dealt adequately with attrition, and that has led a number of researchers to avoid using Project TALENT. In fact, however, Project TALENT staff have dealt adequately

with attrition, and, as one respondent suggested, it is imperative to make that fact widely known. In a recent presentation to the American Educational Research Association (Wise, 1977), the deputy director of Project TALENT made a forceful argument that the TALENT subject-weighting scheme virtually eliminates nonresponse bias. This message should be repeated frequently to counteract a prevalent misconception.

The third step in the process of becoming a data bank user concerns the objective evaluation of the appropriateness of the data for one's research problem. There appears to be a substantial barrier to use of the data when detailed information on the data base is not directly in the researcher's grasp. In response to this need, the Project TALENT Data Bank Handbook has been produced and updated. That Handbook aims to convey all the information necessary to specify exactly what the data base contains. The necessity of charging a substantial amount for the Handbook, rather than being able to mail it free to those expressing interest in use of TALENT data, is a continuing problem that could be solved with a small investment. A subsidy by NIE or AIR for the printing and mailing of copies of the Handbook would appear to be a prudent tactic. This might amount to $1,000 per year.

Of course, no handbook can supply information in enough detail to tell a potential research user that his/her research plan will be successful using a particular data base: if the topic of the research involves a particular type of individual, the numbers of respondents of that particular type in the data base may be insufficient for the proposed research. One solution to the need for a chance to find out roughly the power of the data base for a particular research problem is to provide a small sample of the data which can be thoroughly explored at small cost. Project TALENT has, in late 1977, created a file of 4,000 respondents to the 11-year follow-up survey to be used for such exploratory analyses. The respondents on the Exploratory Tape were selected so that no differential weighting would be necessary and so that the tape would, in general, be easy and inexpensive to use. It is the intent of Project TALENT to provide a copy of this file to qualified researchers for quick and inexpensive analyses for a fee that covers the reproduction of the file and documentation.

The fourth and final hurdle to be overcome in persuading a potential data bank user to become a data bank user is the obtaining of funds to pay for data transfer and/or analyses. The findings of the present survey suggest that a wide variety of perspectives exist with regard to funding. Conversations with representatives of funding sources outside NIE indicated neutrality and general lack of knowledge about Project TALENT as a data base for secondary analyses. While Project TALENT was not being mentioned explicitly in RFPs or research program announcements, neither was there a prejudice against research conducted using Project TALENT data.

The only obvious solution to this hurdle is to reduce costs to the potential user, either through greater efficiency or through subsidy. In view of the level of charges for projects in the recent past, there is no reason to believe that AIR could contribute substantially to reducing costs, other than in the ways mentioned above. Subsidy by a funding source does not appear viable either, except in terms of a favorable attitude concerning research proposals that aim to perform secondary analyses using Project TALENT data. Presentations should be made by Project TALENT representatives to assure that favorable attitude.

Aside from these activities to deal with the particular barriers, one valuable tactic emerged from this survey: the mere conduct of the survey, writing and talking to leading researchers in relevant areas, appeared to substantially increase their personal interest in using Project TALENT data. Because of funding lags, the effect of this activity in terms of data bank usage may not be felt for a year or more. On the basis of initial indicators, however, it would appear to be important to assign a small amount of effort to continued perusal of the relevant journals and selective contacting and interviewing of appropriate researchers.

Bibliography

Babbie, Earl R. Survey Research Methods. Belmont, Calif.: Wadsworth, 1973.

Bell, Charles G. The joys and sorrows of secondary data use. In Bisco, Ralph L. (Ed.), Data Bases, Computers, and the Social Sciences. New York: John Wiley & Sons, 1970.

Borman, Lorraine, & Haig, Richard. The information transfer process on the university campus: The case for public use files. Drexel Library Quarterly, 13(1), January 1977.

Carroll, Steven J., & Morrison, Peter A. National Longitudinal Study of High School Seniors: An Agenda for Policy Research. Santa Monica, Calif.: Rand Corporation, June 1976.

Center for Human Resource Research. National Longitudinal Surveys Handbook. Columbus: Ohio State University, 1974.

Chinitz, Benjamin. The management of federal expenditures for research on social problems. In Horowitz, Irving Louis (Ed.), The Use and Abuse of Social Science. New Jersey: Transaction, Inc., 1971.

Ennis, P. H. Data galaxies in the social sciences. American Behavioral Scientist, VII(1), June 1964.

Flanagan, John C., et al. The American High School Student. Pittsburgh: American Institutes for Research, 1964.

Flanagan, John C., et al. The Career Data Book: Results from Project TALENT's Five-Year Follow-Up Study. Palo Alto, Calif.: American Institutes for Research, 1973.

Flanagan, John C., et al. Five Years after High School. Palo Alto, Calif.: American Institutes for Research, 1971.

Glass, Gene V. Primary, secondary, and meta-analysis of research. Educational Researcher, 1976, 5, 3-8.

Hyman, Herbert H. Secondary Analysis of Sample Surveys: Principles, Procedures and Potentialities. New York: John Wiley & Sons, 1972.

Jencks, Christopher, et al. Inequality. New York: Basic Books, 1972.

Koehler, Virginia. Secondary analysis from a Federal perspective. Paper presented at the annual meeting of American Education Research Association, April 1977.

Levinsohn, Jay, Riccoboro, John A., & Moore, R. Paul. National Longitudinal Study of the High School Class of 1972: Base-Year and First-Year User's Manual. North Carolina: Research Triangle Institute, April 1975.

Nasatir, David. Data Archives for the Social Sciences: Purposes, Operations and Problems. Paris: UNESCO Publications Center, 1973.

Sessions, Vivian S. Primary and secondary data base professionals: Time for rapprochement? Review of Public Data Use, 3(1), January 1975.

Turnbull, William W. (Chairman). Population Changes and the SAT Score Decline, 1960-1972: A symposium at the 1977 annual meeting of the American Psychological Association, San Francisco.

Wilson, Sandra Reitz, & Wise, Lauress W. The American Citizen: 11 Years after High School. Palo Alto, Calif.: American Institutes for Research, 1975.

Wise, Lauress L., McLaughlin, Donald H., & Steel, Lauri. The Project TALENT Data Bank Handbook. Palo Alto, Calif.: American Institutes for Research, March 1977.

Wise, Lauress L. The fight against attrition in longitudinal research. Paper presented at the April 1977 meeting of the American Educational Research Association, 1977. (ERIC Document 137 355)

APPENDIX A

DATA USER INTERVIEW GUIDE

A. Review available information on each user (read journal article, note positions held, previous articles, etc.)

B. Phone each user, based on best telephone number available.

1. If better phone number is needed, try to obtain it and phone again.
2. If address has changed, get it.
3. If user is not available, ask for best contact time or set appointment to call.

C. Conduct interview. Have brochure available for reference.

Introduction: *Hello, my name is _____ and I am calling from Palo Alto, California, on behalf of Project TALENT, a Federally supported data base. By now you should have received a brochure in the mail about Project TALENT. Has it arrived?*

[If yes, go to 1.]

[If no, say this and go to 1:]

We are conducting a survey of professional researchers to find out what factors will facilitate or inhibit researchers from using data bases like Project TALENT. We would appreciate your response to several short questions.

Project TALENT started collecting information on 375,000 high school students in 1960 to study their life development. Data included detailed information on their cognitive skills, interests, plans, family backgrounds, and current activities. These individuals were contacted again at ages 19, 23, and 29 to obtain longitudinal data on their educational, occupational, and social experiences. A fourth follow-up survey of respondents at age 35 is scheduled to begin this fall.

(1) *Have you heard about Project TALENT before? How?*

(2) *What is your current research interest?*
Are you (also) interested in education, family, or career development?

(3) *What barriers do you see that would prevent a researcher such as yourself from using data bases like Project TALENT?*

[Probes: Chances of funding
Availability of key variables
Costs
Unfamiliarity with TALENT
Use of someone else's data]

(4) *What are your recommendations for increasing the use of data bases like Project TALENT?*

(5) *We are now formulating the questionnaire for our follow-up of respondents at age 35. What topical areas do you feel we should address?*

That completes my list of questions. Thank you very much for your time and cooperation.

APPENDIX B

FUNDING SOURCE INTERVIEW GUIDE

A. Review available information on each contact (read journal article, note positions held, previous articles, etc.)

B. Phone each contact, based on best telephone number available.

1. If better phone number is needed, try to obtain it and phone again.
2. If address has changed, get it.
3. If contact is not available, ask for best contact time or set appointment to call.

C. Conduct interview. Have brochure available for reference.

Introduction: *Hello, my name is _____ and I am calling from Palo Alto, California, on behalf of Project TALENT, a Federally supported data base. By now you should have received a brochure in the mail about Project TALENT. Has it arrived?*

[If yes, go to 1.]

[If no, say this and go to 1:]

We are conducting an informal survey of the federal research system to find out what factors will facilitate or inhibit researchers from using data bases like Project TALENT. We would appreciate your response to several short questions.

Project TALENT started collecting information on 375,000 high school students in 1960 to study their life development. Data included detailed information on their cognitive skills, interests, plans, family backgrounds, and current activities. These individuals were contacted again at ages 19, 23, and 29 to obtain longitudinal data on their educational, occupational, and social experiences. A fourth follow-up survey of respondents at age 35 is scheduled to begin this fall.

(1) *Have you heard about Project TALENT before? How?*

(2) *Does your agency/organization fund research studies utilizing existing data bases like Project TALENT?*

(3) *What barriers do you see that would prevent your agency from sponsoring projects using Project TALENT data?*

[Probes: Chances for funding
Availability of key variables
Costs
Unfamiliarity]

(4) *What recommendations would you make for us to increase the use of TALENT data?* [Probe for dissemination strategies, etc.]

Thank you very much for your time and cooperation.